PRACTICAL RISK ADJUSTMENT FOR PAYING HEALTH CARE PROVIDERS – IS THE CREAM THAT IS SKIMMED CREDIBLE OR MUST WE SHRINK FROM NORMAL DEVIATIONS?

James F. Burgess, Jr. Cindy L. Christiansen Willard G. Manning

Carl N. Morris*

July 29, 2002

*Conclusions and opinions expressed herein are those of the authors and do not necessarily reflect official positions of the U. S. Department of Veterans Affairs (VA). Corresponding author is James F. Burgess, Jr., Management Science Group, Department of Veterans Affairs (518/MSG), 200 Springs Road, Bedford, MA 01730, USA. Tel.:+1 – 781 – 687 – 2488; fax: +1 – 781 – 687 – 2376; e-mail: burgess@world.std.com. Burgess also is affiliated with the Boston University School of Public Health; Christiansen is affiliated with the Boston University School of Public Health and the VA Center on Health Quality, Outcomes, and Economic Research; Manning is affiliated with the University of Chicago and acknowledges funding support from the VA Management Science Group; and Morris is in the Statistics Department at Harvard University and acknowledges funding support from the VA Center on Health Quality, Outcomes, and Economic Research.

Abstract

Keywords: Risk adjustment; Resource allocation; General linear models; Loss functions; Hierarchical models; Regression-to-the-mean;

1 Introduction

When health plans or government agencies are paying provider groups, what should the role of risk adjustment be and what difficulties are encountered when diagnostic based risk adjustment is employed. Historically, relatively simple age-sex adjustments have represented the extent of risk adjustment for providers; however, Newhouse has emphasized the important economic tradeoffs theoretically (Newhouse, 1995) and practically (Newhouse, 1998) entailing considerations of more extensive approaches. We attempt to provide an intuitive and practical discussion of this aspect of the problem of paying provider groups a sufficient amount to cover average standards of care across diagnosed diseases. This problem is faced generically by health plans, state governments, the federal government in the US, plus regional health ministries in other countries. But first, we need to be clearer about what we mean by risk adjustment.

Any discussion of risk adjustment must begin with a clear description of the particular definition of risk adjustment and problem under consideration since the concept itself is well worn but not universally defined (Dowd and Feldman, 2001). We begin with the common starting point of postulating a functional relationship describing health care expenditures for individual i over a given period, such as a year: $\text{Expenditures}_i = f(X_i, u_i)$. In our case, the vector of observed explanatory variables X primarily will be diagnosis based indicator variables from a commonly used commercial risk adjustment system (the Diagnostic Cost Group/Hierarchical Condition Category (DCG/HCC) model recently chosen as the risk adjustment tool for use in Medicare+Choice).

Some of the statistical approaches to explaining health care expenditures have tended to fall back on simple ordinary least squares (OLS) regression models as the functional form of $f(X_i, u_i)$ (see Blough et al. (1999) and Duan, Manning, Morris, and Newhouse (1983) for some exceptions). This OLS modeling focus takes place in an environment where stable

and unobserved latent factors in populations exist that can never be fully accounted for by managers, modelers, or researchers (Newhouse, 1998). The distribution of residuals in such regressions offer potential opportunities for cream skimming off these latent factors either by health services providers (e. g. by choosing services to focus on or de-emphasize) or by insurers (e. g. in setting barriers to particular patient populations). Stability and unobservability from the perspective of the modeler represent necessary conditions for the type of cream skimming we postulate since random outcomes/behavior in a single year or factors that can be observed by the modeler can be accounted for. So, to summarize, risk adjustment in the context of the present paper is a process of using clinical and possibly utilization information to calculate the expected health expenditures of individual consumers in a health care provider system over a year to allocate resources to particular provider groups within the system. Our discussion has wider implications as well, that we will attempt to illuminate as we progress.

We discuss regression model approaches in the functional form that can address one specific type of opportunity for cream skimming. We focus on what might happen when one can identify specific types or groups of patients who are systematically mispredicted – both in estimation samples and prediction samples – that cannot or are not accounted for in regression. Regression analysis also assumes a particular symmetric loss function formulation that tends to be accepted without critical consideration in going from estimation samples to prediction samples (specifically mean squared error when discussing OLS). Next, we discuss the implications of symmetric loss functions in the context of common issues and purposes for explaining health care expenditures in populations. The sense of the loss function we discuss here is totally statistical, and does not account for economic deadweight losses that one also might calculate in considering the behavior responses to cream skimming incentives. Finally, we also discuss hierarchical model approaches that could be

employed, regardless of which functional form approach is used, that would control the degree of credibility we put in individual level risk adjustment for particular characteristic differences in the underlying patient populations.

One could approach these issues in a highly technical statistical way; however, this paper attempts to take an intuitive focus aimed at a wider readership. An intuitive approach requires a keen focus with good generalizability to a wide range of applications. The focus of the paper throughout will be on a process for approaching these problems in risk adjustment in the context of a specific example of resource allocation.

2 A Resource Allocation Example

In practice, there are many ways that risk adjustment could be brought to bear to address problems in resource allocation once a potential or actual treatment population is identified. Common examples include the case of health plans devising capitated contracts with providers or groups of providers, staff model health maintenance organizations allocating resources to subgroups of providers, self insured employers allocating resources for employee health coverage across geographically separated plant sites, or countries with nationalized health insurance/providers allocating funds to localized trusts. The main goal is to achieve fairness in the allocation of fixed (or quasi-fixed if enrollment is open) assets to cover health care services. By fairness in this context, we mean that funds are provided such that sufficient resources are available with efficient management to treat patients to the community standard of care. Careful balancing to achieve this is extremely difficult, since objective standards for the provision of health care varies substantially across conditions and incentives facing patients and providers are complex. The focus of this example, though, is on sufficiency for the providers or to a geographic unit to which resources are

being allocated, while recognizing that other standards are possible.

Fairness in this sense and risk to the providers are not precisely mirror images of each other since fairness to the providers only requires mitigation of risks that are out of the control of providers. One of the chief factors that should be out of the control of providers is the profile of the clinical complexity of the patients. While debates may rage over the potential effectiveness of cream skimming by providers in avoiding the most clinically complex patients, an effective solution is to mitigate those risks, and thus the incentive for cream skimming, directly. Recently, measuring the clinical complexity of patient populations to discourage cream skimming has generated a range of commercially and academically developed products with different characteristics, but again these discussions generalize to any of these systems and models using those systems. These products share a focus on diagnoses and grouping of diagnoses to classify patient risk at the capitated level. These indicators can then be used as some of the independent variables X_i in a regression analysis on patient level expenditures. Resources can be allocated by payments/allocations based on predicted expenditures from these regressions.

As mentioned above, in this paper we will be using Version 5 of the DCG/HCC model in this role. In particular for the HCC-DCG models, but for most of these risk adjustment models in general, OLS models without intercepts are used by the developers of the methods (Ellis et al. (1996); Ash, Ellis, at al. (1998); and Pope et al. (1998)). A number of statistical and economic issues are raised/addressed by this choice of a model structure (also discussed by Van de Ven and Ellis (2000)). One important factor is having monotonicity in estimated predicted costs in incremental increases in risk adjustment characteristics. By selecting models with only positive parameters, this means that any additional diagnosis in a patient record brings a positive incremental predicted increase in expenditures - models with intercepts might not have this property. Another important factor is simplicity or

ease of explanation to policy makers or managers, despite potential difficulties even in truly understanding OLS regression. What managers and policymakers actually seem to value is the ability to build a model into a spreadsheet for ease of use and the capability of making intuitive explanations of the implications of models – characteristics not limited to OLS estimates.

One of the reasons that OLS models have been used is difficulties with estimating in samples of 1,000,000 or more with alternative functional forms, sample sizes that are common in the DCG model development work cited above. But such 1 million/3 million/10 million patient samples can still create cells per HCC per geographical unit that can be quite small for rare conditions. Obviously, there are balancing acts between unwieldy models with too much data and these same complicated models with less data where these small cell problems can be even greater. This variation in statistical properties across cells suggests some type of hierarchical or multi-level model that focuses credibility issues on these smaller cells. Another important difficulty with more complicated models that varies with the type of complexity is the retransformation problem (Manning and Mullahy (2001)). And there also is difficulty in managing the attempt to avoid overfitting problems, especially in out of sample prediction, in practice. The standard best linear unbiased estimator concepts of OLS also are important; however, using models with suppressed intercepts leads to biased estimates.

One major issue not addressed by these issues is skewness in the distribution of the underlying health care expenditures at the individual patient level. The skewness of use of health care services in general is well known (Duan, Manning, Morris, and Newhouse (1983)), summarized by Jones (2000), and drawing on Manning (1998) and Mullahy (1998). Much of that discussion and focus has been on the log transformation and the difficulties involved in retransformation (further expanded on by Manning and Mullahy (2001)). The

log transformation is popular, but could overcorrect or undercorrect for the underlying skewness in the data depending on the context. As we will develop below, a square root transformation clearly outperforms the log transformation in the particular health expenditure data example that we employ. In particular, skewed data can lead to situations where the predicted payment value is higher then actual costs for most patients, while high cost patients are still severely underpredicted. Other combinations are possible as well from different types of skewness or kurtosis, but intuitively it seems to be related to combining distributions of illness and health together, aggregating at the patient level. Explaining such issues to health care managers and policymakers is difficult, especially when even complex econometric techniques do not solve the problem easily.[1]

Another thing heretofore undiscussed and unrecognized as an issue in risk adjustment, complicated by data skewness, is the implicit assumptions about the nature of the loss function imposed by regression. One of the goals of this paper is to evaluate the implications of the symmetric loss function against the fairness and incentive goals for this type of risk adjustment. In regression, especially OLS regression, the loss function is symmetric in the units of the dependent variable, which may be transformed. Therefore, an evaluation needs to take account of the implicit changes in the loss function as different functional forms (e. g. log, square root, gamma) are employed. More importantly, one may not desire an equal treatment of overprediction (patients whose predicted expenditures are greater than their actual expenditures) and underprediction (patients whose predicted expenditures are

[1]We looked carefully at both Box-Cox and GLM models among these more complicated models. The Box-Cox error terms on the transformed estimation scale indicated a type of complex form of heteroskedasticity that ruled out a direct retransformation fix (Manning, 1998). The GLM model allows the skewness to be addressed by the choice of a distributional family (in our case a Gamma) matching the mean/variance relationship, while the non-linearity is separately addressed by the power choice in the link (in our case a square root).

less than their actual expenditures).

The cream skimming incentive reliably focuses increased relative attention on underprediction. If providers or health plans also can reliably identify the stable part of the distribution of patients who are more likely to be underpredicted, using information not readily added to the risk adjustment process, then the potential for cream skimming exists (Newhouse (1998)). This relationship might not be completely uniform, depending on the potential nature of these unmeasured factors across patients. Nevertheless, a properly designed asymmetric loss function could balance these incentives more appropriately with relative ease.

Assessing the appropriateness of resource allocation is more complex as it could affect the choice of a loss function. One requirement would seem to be that the mean payments or resource allocation at a provider group level not be skewed by any change in the risk adjustment loss function. Such an outcome could not be generalized, but must by definition be context specific, though the allowance of individual deviances, but adherence to group means, could be written into an objective function. Another aspect of the desired goal is to provide uniform incentives for quality of care across all diseases. Though we employ disease specific risk adjustment, we do this at the mean level, not the variance level, and individual diseases have different ranges of potential treatment costs and different levels of quality of care concerns. The loss function could incorporate these differences as well, such that asymmetry to allow underpayment is more restricted when treatment costs (and thus courses of treatment) and quality are more highly variable. Addressing this latter issue is beyond the scope of this paper; however, it is important to keep it in mind throughout.

Clarifying all of these issues requires the selection of good examples which illustrate the range of concerns expressed. We choose to use the Department of Veterans Affairs (VA) health care system as such an example. For more than five years, VA has been allocating

national Congressionally appropriated resources to 22 (now 21 as two smaller ones have been merged) Veterans Integrated Service Networks (VISNs) to deliver local care to eligible veterans. The nature of this problem is quite similar to the problem of an HMO capitating patient care to a set of provider groups or other similar problems outside of the government setting. It also provides a large amount of data to set up estimation and prediction samples so that out of sample prediction and overfitting problems may be studied. In addition, VA is considering DCG/HCC risk adjustment for use in their VISN resource allocation system.

2.1 Specifics of VA Example

The VA operates the largest health care system in the US with 163 hospitals, more than 800 community and facility-based clinics, 135 nursing homes, and other facilities. With a medical care budget of more than $19 billion in FY2000, out of the total population of 26.4 million veterans, VA provided care to 3.8 million unique users, 3,000,499 of whom were provided care under priority for service connected disabilities, meeting an income/wealth based means test, or from a variety of smaller health care need and veteran specific reasons. 2,979,760 of these patients have measured costs accurate enough to be included in the patient sample that serves as the sampling population for our analysis.

This FY2000 dataset of the universe of VA Congressionally reimbursed users, consisting of 2,979,760 observations, was randomly sampled into two groups of 100,000 observations each. One sample is the estimation sample and the other forms a prediction sample that will allow testing for overfitting sensitivity. An advantage of the VA population is that it is large enough itself to make these relatively large subpopulations from a sufficiently larger base. Note that most practical public or private populations in managed care plans or health care provider systems in the US fall in the range of this sample size (see Dunn, 1998, for example, which analyzes risk adjustments in four populations of 240,000, 120,000, 115,000

and 70,000). Many people also test risk adjustment models on large scale health care surveys that collect and link claims data and diagnoses, such as the Medical Expenditure Panel Survey and the Medicare Current Beneficiary Survey with much smaller sample sizes.

In recent years, the most important advances in risk adjusting patient populations to explain health care expenditures have employed diagnostic information to characterize disease patterns. There are two basic strands of analysis that flow from this work. Most commonly, analysis has focused on predicting the health care utilization of enrolled patient populations next year from diagnoses and other information (possibly even including costs) collected this year. This is called prospective modeling. However, a growing application of risk adjustment is in helping integrated health care delivery systems or insurers understand differences in the risk of current populations, for budget allocation or rate setting purposes. We employ this type of concurrent modeling in this paper. For the VA setting there is a large pool of potential patients (26 million) and a persistent, but not identical, pool of year to year users (3 million, though this is growing at present as long as alternatives in the private sector offer restricted benefit packages [e. g. Medicare with outpatient prescription drugs]). As a result, without explicitly enrolled patients, the VA does resource allocation to the VISNs on a rolling two year regulatory lag. For example, the FY2000 data used here was used by VA to build the FY2002 VISN payments or budgets – though a diagnosis based risk adjustment system has not been adopted there at this time. The intuition for this particular example follows from that context.

To characterize the explainable portion of variation in expenditures, we begin with Diagnostic Cost Group (DCG)/Hierarchical Coexisting Conditions (HCC) models (Ellis, et al. (1996), Ash, et al. (1998), Pope, et al. (1998)) to group ICD-9-CM diagnoses into HCC indicator groups as explanatory variables for health care expenditures. This model takes the 15,000 ICD-9-CM codes, groups them into categories and then places the

groups into body system/clinical condition specific hierarchies. These hierarchies allow some patients to have multiple HCC's and disallow others, helping to address overfitting problems when people with complex diagnoses also by definition have less complex ones in the same hierarchy. Out of the 118 HCC's in Version 5 of the DCG model, we employ 69 HCC's in our model that are the most statistically significant in explaining differences in expenditures for the VA population. HCC's are not mutually exclusive, veterans can and do have many more than one HCC, in fact, as we will see it is the veterans with more than 10 HCC's who are particularly problematic to estimate. Brief descriptions of these HCC's are reported in Table 1.

In addition, VA has a number of special emphasis programs and offers care for a wide range of disabling conditions that add to expenditures for veterans served by those programs or having those disabling conditions. VA has established systems for defining membership in these groups and they are assigned hierarchically, according to increasing expected costs, so that no veteran can appear in more than one special population group and they appear in the highest cost group for which they qualify. Fourteen of those groups are used in this analysis, identified by brief descriptions in Table 1 as well.

In addition, the set of explanatory variables X includes a gender identifier and age variables. First, the age variable is centered at 60, near the mean age for VA patients. But some ages for veterans are missing or have implausible values (e. g. below 20) and these are replaced with the mean age before centering. Then these observations are identified with an agemiss variable that also is used in the analysis. Finally, the centered age variable is squared and that variable is used as well.

TABLE 1: HCC and Special Population Flags	
Name	Description
HCC flag ID	HCC description
hcc001	hiv/aids
hcc002	septicemia (blood poisoning)/shock
hcc003	central nervous system infections
hcc004	other infectious disease
hcc005	metastatic cancer
hcc006	high cost cancer
hcc007	moderate cost cancer
hcc008	low cost cancers/tumors
hcc013	diabetes with chronic complications
hcc014	diabetes with acute complications
hcc015	diabetes with no or unspecified complications
hcc016	protein-calorie malnutrition
hcc017	moderate cost endo/metab/fluid-electrolytes
hcc019	liver disease
hcc020	high cost chronic gastrointestinal
hcc021	high cost acute gastrointestinal
hcc022	moderate cost gastrointestinal
hcc023	low cost gastrointestinal
hcc024	bone/joint infections/necrosis
hcc025	rheumatoid arthritis/connective tissue
hcc027	aplastic and acquired hemolytic anemia
hcc028	blood/immune disorders
hcc029	iron deficiency and other anemias
hcc030	dementia
hcc031	drug/alcohol dependence/psychoses
hcc032	psychosis/higher cost mental
hcc033	depression/moderate cost mental
hcc040	quadriplegia
hcc041	paraplegia
hcc042	high cost neurological
hcc043	moderate cost neurological
hcc044	low cost neurological
hcc045	respirator dependence/tracheostomy
hcc046	respiratory arrest
hcc047	cardio-respiratory failure and shock

TABLE 1: HCC and Special Population Flags cont.	
Name	Description
hcc048	congestive heart failure
hcc049	heart arrhythmia
hcc050	acute myocardial infarction
hcc051	other acute ischemic heart disease
hcc053	valvular and rheumatic heart disease
hcc054	hypertensive heart disease
hcc055	other heart diagnoses
hcc058	high cost cerebrovascular disease
hcc059	low cost cerebrovascular disease
hcc060	high cost vascular disease
hcc061	thromboembolic vascular disease
hcc063	other circulatory disease
hcc064	chronic obstructive pulmonary disease
hcc065	high cost pneumonia
hcc066	moderate cost pneumonia
hcc067	low cost pneumonia
hcc068	pulmonary fibrosis/other chron lung
hcc069	pleural effusion/pneumothorax
hcc075	low cost ear, nose, and throat
hcc076	dialysis status
hcc077	kidney transplant status
hcc078	renal failure
hcc079	nephritis
hcc080	other urinary system
hcc091	chronic ulcer of skin
hcc093	vertebral fractures/spinal cord injury
hcc094	hip fracture/dislocation
hcc095	head injuries
hcc096	drug pois/intern injur/traum amput
hcc097	other injuries and poisonings
hcc098	complications of care
hcc099	major symptoms
hcc100	minor symptoms, signs, findings
hcc102	high cost congenital/pediatric

TABLE 1: HCC and Special Population Flags cont.	
Name	Description
hcc103	moderate cost congenital
hcc110	major organ transplant status
hcc111	other organ transplant/replacement
hcc112	artificial opening status/attention
hcc113	elective/aftercare
hcc114	radiation therapy
hcc115	chemotherapy
hcc116	rehabilitation
hcc118	history of disease
Special flag ID	Special population flag description
Flag1	Spinal cord injury
Flag2	Chronic mental illness
Flag3	Traumatic brain injury
Flag4	Blind rehabilitation
Flag5	Post-traumatic stress disorder
Flag6	Alcohol and drug addiction
Flag7	Stroke
Flag8	Hepatitis C
Flag9	Home health care
Flag10	Domicilliary
Flag11	Long term care
Flag12	AIDS
Flag13	Transplant
Flag14	ESRD

The general structure of budgeting in the VA system is beyond the scope of this paper (the general issues are presented in Lehner, Burgess, Hults, and Stefos (1996)); however, a few salient facts must be presented for context. Much Congressional scrutiny is focused on the fairness of these distributions of appropriated resources to local levels since these represent state and congressional district federal spending (reference GAO study and RAND study). The current allocation system only has three prices, from grouping patients into three very large groups, which provides very little disease specific risk adjustment. Both of these studies urge some wider, more specific way of categorizing patients, though

DCG/HCC risk adjustment is only one possibility under consideration.

Here we step aside from the specific VA problem to postulate a related one, that illustrates our problem better. The importance of risk adjustment increases as the sample size decreases, but conversely doing the risk adjustment properly gets more difficult if the groups are not experience rated. We have chosen a sample size of 100,000 capitated patients for illustration here. This accentuates problems that might occur only when sample sizes of particular diagnoses drop and creates risks of overfitting. We will test this with cross validation across the randomly selected 100,000 patient populations pulled from the much larger VA population.

We spent a great deal of time and effort evaluating different models and functional forms for $\text{Expenditures}_i = f(X_i, u_i)$, including transformations of Expenditures_i. We began with OLS regression as a basis for comparison. Instead of presenting this wide variety of potential models directly, we will do all of our comparisons to one alternative in the General Linear Model (GLM) family with Gamma distributed errors and a power of 0.5 (square root) link function. In doing this, we will emphasize the process that was used to select this alternative model and some of the important comparisons that we did, rather than the technical statistical methods of how to do this. STATA 7.0 was used to perform these analyses.

3 Description of the Base Models and Results

A very small number of observations in the two randomly selected 100,000 patient samples (three cases in the estimation sample and two in the prediction sample) had missing gender and were deleted from all analyses. The summary statistics presented in Table 2 illustrate the high degree of variability (coefficients of variation near 3) and skewness to the right.

TABLE 2: Dependent Variable Statistics for the Two Samples							
Sample	Mean	Stdev	Skewness	Kurtosis	Median	95%	99%
Estimation	5369	14929	9.111	144	1532	22805	70279
Prediction	5468	15689	12.55	416	1553	23165	71811

For the OLS model, we present the model with an intercept, unlike the no-intercept models used by the DCG model developers (Ellis et al. (1996); Ash, Ellis, at al. (1998); and Pope et al. (1998)), although we ran both. The OLS models have both advantages and disadvantages. The advantages include simplicity and ease of estimation, especially for very large samples. The disadvantages include failure to deal with the skewness in the data, the possibility of negative predictions (when employing an intercept), or bias (for the case of an added flag for no HCC, but a suppressed intercept).[2]

For the GLM model, we considered various power transformations for the link, including the log link, as well as a failed attempt at the identity link. We judged the alternatives not in terms of their deviance, a traditional approach, but in terms of their goodness of fit to the data as measured by two sets of tests. The first of these is Pregibon's Link Test, which was developed for dichotomous models, but can be easily adapted to any GLM. This is a two-step test conducted on the scale of the link function, with the first step estimating the coefficients of the full model and then creating a summary index function on the estimation scale. In the second step, the GLM estimates the response of cost to this index and its square. If the model is linear under the link function, then the coefficient of the squared term should be insignificant.

The second test is a direct test on the scale of interest - actual dollars – or the raw scale. The test is a variant of one suggested by Hosmer and Lemeshow (2000) for checking the fit of logistic models, but can be applied more widely. After estimating the full model, we

[2]We discussed the bias in the no intercept model briefly above, recall that the suppressed intercept is used primarily to ensure monotonicity and no negative predictions in practical applications. It is an open question whether this gain is balanced by the bias it creates.

predict mean expenditures (conditional on the covariates) for each case on the raw scale. We create raw-scale residuals and then a set of flags for prediction categories (twenty of them in five percent ranges, based on prediction). We regress the raw-scale residual on these 20 indicators, suppressing the intercept, and using robust (Huber/White) estimators of the variance-covariance matrix. If the link function is correct and the specification of the X's is appropriate, then the means by subgroups should not exhibit any systematic U-shaped pattern. And the estimates should not be significantly different from zero, using an F test. Given the sample size here, there could be significant but small coefficients. We also used this test to check the fit by age and number of HCC's for all analyses.

Our focus on these tests in general and for age/HCC subgroups begins our investigation of the impact of loss functions on model choice. Deviance based measures, especially squared deviation based measures, line up directly with the symmetric OLS loss function. The Hosmer/Lemeshow and Pregibon tests focus us more generally on avoiding systematic mis-payment for any of the tested subgroups of patients.

These analyses suggested that the optimal power link was about +0.6, though we use the pure square root (+0.5) for simplicity of retransformation. Both the linear and log links led to systematic misfitting in the center of the predicted range, with the direction of the misfit the opposite in the two extremes. The square root link seems to be a reasonable compromise that maintains the linearity of the index function, while fitting the raw-scale relatively well. Moreover, values for the largest Cook's Distance indicate that both the OLS and the log models have more influential observations than does the square root model.

Thus, the functional form of our model assumes that $[E(\text{Expenditures}_i|X_i)]^{0.5} = X_i$ or that, rearranging, $[E(\text{Expenditures}_i|X_i)] = X_i^2$, where the X_i's include the various HCC and subpopulation flags, and demographic variables as described above. Another worry with the square root link is that negative coefficients on the HCC terms will cause large

negative terms in $\hat{X}$ to become large positive predictions, an order of predictions problem that does not arise with identity and log links. Fortunately, in our case, this did not happen. All of the HCC variables had positive coefficients, and the only negative coefficients were for demographic variables.

The other choice in the GLM approach is to select a distribution family. To do this, we need the relationship between the Var(Expenditures$_i|X_i$) and the E(Expenditures$_i|X_i$). We employed the modified Park test (Park, 1966) described in Manning and Mullahy (2001). GLM models expect that the variance function is some power of the mean function, both conditional on X. In this case, we regress the log (the raw-scale residual squared) on the log of the raw-scale prediction. If the slope is near two, the data are typically modeled with a gamma distribution. If near one, we could use a Poisson-like approach. For most of what we examined the estimated slopes were very close to two, indicating a model with a constant coefficient of variation and the choice of a gamma distribution. We did notice that the model was not as well-approximated by the gamma over the range of predictions or the number of HCC's. However, as long as the link and X's are specified well, any problems in the distribution using this part of the information only lead to efficiency losses if the sole concern is getting a payment system that is not systematically biased for some subgroups of patients. When there also is a goal of understanding the distribution of actual costs – to be able to say something about how many cases will cost more than $100,000, or to simulate the distribution of VISN costs – then one would need to find a distribution that more closely approximates the empirical one. We discuss both finding better distributions and implications further below.

3.1 Basic Results for the OLS and GLM Models and Model Checking

Regression results for the GLM and OLS models on the full estimation sample may be found in Tables 3 and 3A respectively in the appendix. Note in Table 3 that in addition to presenting the direct regression results, we present the squared raw scale coefficients in dollars, just as one would present OLS results to policymakers or managers. Nearly all of the parameter estimates are statistically significant and in the raw scale, the smallest increment for practical significance is for HCC 068 (pulmonary fibrosis and other chronic lung) with a value of $26, among six parameters under $50. Note in Table 3A that a handful of the parameters are statistically insignificant, though the choice of HCC's for this model came from an analysis of the entire 3 million patient database where all of these HCC's are statistically significant. The constant is very slightly negative (statistically insignificant); however, no parameter estimate is smaller than the $236 estimate for HCC 075 (low cost ear, nose, and throat) and only four of the parameters are under $500.

The modified Hosmer/Lemeshow tests illustrated the most interesting differences between the OLS and the GLM model. Figures 1 and 2 in the Appendix depict the main results for the 20 categories by five percent range of the rank of cases by prediction. Focusing first on the full estimation sample results (in Figure 1), note that the first ten categories (below the median prediction) for the OLS model all have large positive residuals (all highly statistically significant with t-statistics roughly proportional to the size of the coefficient), indicating that the predictions systematically underpay for these cases. Above the median prediction, in the OLS model, the reverse is true and there is systematic overpayment with generally statistically significant negative residuals. However, the very last category, the top five percent of predictions revert back to slight (statistically insignificant, though larger in value) underpayment. As we see from Cook's Distance and influential observation statis-

tics, the OLS model focuses a great deal of attention on this top five percent of predictions, getting them closer to right, at the expense of poorer fitting everywhere else but near the median of predictions. In this VA population, this would provide a greater incentive for seeking out sicker patients at the expense of healthier patients, which might be an informed policy choice, but was hidden by more cursory analysis.

The GLM model, though, performs much better across the range of predictions, as we can see graphed on the same scale in Figure 1. While 17 of the 20 categories had t-statistics on their residuals of over three, only the highest prediction category has a t-statistic over three in the GLM model. With a value of -$3065 on the raw scale and a t-statistic of -5.23, the GLM Gamma/Square Root link model systematically overpays for this group of patients. Nevertheless, overpaying for the most expensively predicted patients is not necessarily a bad outcome from an incentive perspective. The VA and other organizations sometimes consider various kinds of top-coding or other separate treatment of the most expensive cases to deal with this. Yet, as compared to the OLS model, the GLM maintains a much better overall fit, rather than orienting its estimates to predict this most expensive group well, at least in the case of the VA data we employ.

Further exploratory analysis uncovered two additional major issues in estimation. First, a closer look at the underlying causes of the Hosmer/Lemeshow test outcome shows that the difficulty in estimating high predicted cost patients is most closely related to problems in estimating individuals with a large number of HCC flags turned on, in particular ten or more. A coefficient on those individuals with ten or more HCC's for the GLM model has a raw scale value of -$5138 with a t-statistic over seven, a highly problematic overprediction problem. The OLS model on the same test systematically underpredicts by $3116 for these ten or more HCC patients and systematically overpredicts patients with 3-5 HCC's with statistically significant t-statistics over five. We tried to remedy this problem with a wide

variety of *ad hoc* and formal specifications, but did not find one that significantly improves on the problem.

More intuitively, though the nature of the problem differs in the OLS specification as opposed to the GLM specification, when we use the same HCC parameters for all patients then people with multiple HCC's are adding the same amount to the estimation (square root) scale predictions whether or not they have other HCC's (possibly comorbidities that are related) or not. In the GLM, on the square root scale, the incremental cost of adding one more disease appears to be less if the patient already has a large number of HCC's than if they have fewer.

Finally, on the other end of the distribution of the number of HCC's, we had a hugely significant (t-statistic of 42) coefficient of $810 on the raw scale for the GLM on the patients with no HCC's. The basic GLM model significantly underpredicted these patients, and the parameter estimates on their demographic characteristics seemed to be significantly different as well. Though the OLS model predicted these no HCC patients very well, we explored the impact of separating the sample and just estimating the patients with at least one HCC separately. The regression results for the GLM and OLS models on this restricted group of 81,725 patients in the estimation sample may be found in Tables 4 and 4A in the appendix respectively. Figure 2 in the appendix also illustrates the results of the 20 category, ranked by predictions, Hosmer/Lemeshow test for this restricted sample with at least one HCC. For the OLS model, the results are similar, though the coefficient on the five percent highest predictions category is larger ($1406 vs. $912) and more significant (2.07 vs. 1.58). Similarly for the GLM model, the most important difference in the restricted model for this test also is on the five percent highest predictions category (-$4025 vs. -$3065 in coefficients for the at least one HCC sample vs. the full sample).

We will study this restricted sample more completely with respect to general evaluation

and loss functions below. The figures also illustrate the impact of the Hosmer/Lemeshow test statistics on the hold out prediction sample. Here we search in particular for overfitting problems affecting the impact of the parameter estimates on a new sample. The OLS model on the full sample is nearly identical across the range, except again in that 20th category, the 95-100% largest predictions. Here the coefficient goes from $912 to $2058 and the t-statistic increases in the prediction sample to 3.31 and statistical significance. Again, this is a substantially increased underprediction problem that could create a more severe cream skimming incentive. The same pattern appears in the restricted sample of those with at least one HCC. And just as before, these results are strongly associated with a similar incentive to avoid patients with 10 or more HCC's. In the GLM model on the full sample, comparing the Hosmer/Lemeshow statistics between the estimation and prediction sample also shows the most important difference in the 20th category; however, the effect goes the opposite way. The coefficient is smaller (-$3065 to -$2555) and less significant (-5.23 to -3.97) and the same pattern is exhibited in the restricted sample with at least one HCC. The importance of these Hosmer/Lemeshow statistics is underscored in the next section when we look more closely at evaluating forecast procedures.

4 Methods for evaluating forecast procedures: Separation of fitting and forecast methods; and Loss functions

4.1 Data for modeling fitting (F-data)

As we just have seen in the preceding section, statistical modeling and evaluation can be, and should be, separated. The model fitting is conducted on a data set denoted "F" (for

fitting in this section, what we called the estimation sample above) used to reveal the probabilistic relationship between outcomes (in our examples, annual expenditures of patients) and of selected predictors, or covariates, or combinations thereof (diagnoses, severity flags, etc, in our example). Selecting an appropriate model does not require specifying a measure of prediction inaccuracies (via a loss function).

Forecast illustrations in this paper are conditional on knowing a specified case-mix of predictors (so the number of patients in the future at each facility and the percentages in each diagnosis group are known, e.g.). In practice, forecasts of other quantities also would be required, but while uncertainty about these matters would add to overall uncertainty, the statistical modeling issues that arise in more comprehensive situations do not differ fundamentally from what we show here. The fitting results illustrated here require estimating a probability distribution for each combination of individual covariates (demographic combinations, diagnoses, severity flags, etc.). Assuming individuals are independent, at least within medical groupings, these results provide a joint probability distribution for individual outcomes (annual expenditures) as dependent on specified predictors.

4.2 Data for evaluations (E-data) of forecasts

Evaluations of forecasts also are done with a data set ("E" for evaluation, in this section, what we called the prediction sample above). Too often, the data sets F and E are the same, but that provides unconvincing evaluations. A new data set is better for doing evaluations, and in our case the VA data set has been split into 2 equal parts, F and E, each with about 100,000 VA patients (both randomly drawn from nearly 4 million patients treated in 2000).

4.3 Loss functions

Evaluations then depend on how closely outcomes in the evaluation (E) data can be predicted from predictors in the fitting (F) data, based on the prediction model developed with E data. "Closely" will be taken to mean that differences, or errors, depend on differences between predicted outcomes and actual outcomes. Differences are measured by a loss function (negative of a utility function) that involves both data sets (F and E). The loss functions we have considered all take the form:

$$\text{Loss} = \sum_j (W_j * D(\text{error}_j)), \tag{1}$$

where the error is defined as:

$$\text{error}_j = \text{prediction}_j - \text{outcome}_j \tag{2}$$

Sums in this loss function are over j in a set J, where J could be all individuals in F, or it could be an index set for well-defined clusters of individuals (e.g. VISN groupings, or demographic, or disease type, or combinations thereof), for the forecast data set F. The function $D()$ is 0 when an error $= 0$ and D increases monotonically in both directions as errors move away from 0. The weights $W_j > 0$, which are permitted to depend on outcomes, specify the relative importances of individuals or groups, perhaps reflecting the number of members in group j.

The choice of loss function can greatly affect which estimates are optimal and the Loss (total). Common and natural choices for $D()$ include squared errors and absolute errors, for which the optimal estimates (minimizing expected Loss) are the mean and the median, respectively. For symmetric data, e.g. if the outcomes follow Normal distributions, these are the same. But if outcomes are noticeably skewed, as are individual health expenditures

in most data, including ours, then the mean will be substantially larger than the median (E. g. the mean is more than three times the median for individual expenditure VA data – the Gamma distribution with convolution parameter equalling 0.72), so choosing absolute error Loss when working with a skewed expenditure distribution strongly encourages underestimates of average individual expenditures. Aggregate expenditure then will be heavily underestimated.

However, Losses will be asymmetric when allocating budgets to administrative units because underestimates foster perverse incentives (e. g. creaming and dumping) that degrade patient care, while overestimates avoid this hazard. In that case, $D()$ would take larger values for negative errors and smaller values for positive errors of the same magnitude. The non-symmetric loss we have chosen to use is:

$$D(\text{error}) = \max(\text{error}, -k * \text{error}), k > 1. \tag{3}$$

Here, $k = 1$ corresponds to absolute error, for which the median is the optimal estimate. The optimal estimate for skewed loss with an arbitrary value of k is to use the the p-th quantile, with $p = k/(1 + k)$. Thus, $k = 2$ indicates that underestimation errors are twice as costly as those that overestimate. Individual level expenditure distributions at the VA have the mean roughly at the 65th quantile, so the mean for this skewed loss is nearly optimal.

Loss functions that emphasize individual patient estimates are inappropriate for the management policy issues being discussed here. Budgets for particular administrative units depend on getting aggregate totals right, and not forecasts for individual patient components. Because aggregated costs tend to be symmetric and Normally distributed (the central limit theorem), so we emphasize the Normal distribution in our evaluations (E), even though cost data (F) for individuals are correctly modeled as highly skewed.

The weights W_j are chosen to reflect the number of patients in an aggregate. They also can depend on the outcome targets (e.g. the prediction$_j$ from the F-data). The Gamma model fits the individual patient data much better than the Normal distribution, but fitting these distributions requires concentrating much more than OLS on small expenditures because the criterion in this scale family invokes percentage errors, not absolute errors. Although the modeling and evaluation components are separated here, the Gamma of course will gain especially for Loss functions that depend on percentage errors, i. e. on (|error$_j$|/prediction$_j$).

We have not emphasized that there is a natural scale to be estimated here, i. e. the costs, in dollars. Modeling might be pursued in some non-linear form, e.g. someone might use logarithms of costs, and our square-root link Gamma model relates most immediately to the square-root of costs. That may be appropriate for modeling, but for the evaluation, retransformation back to raw dollars must be undertaken, and that can be hazardous. See Duan (1983) for the related issue with respect to smearing estimates. When dealing with administrative budgets, this "smearing" issue simply means getting the total budget right – the sum of all forecasts has to add up to a total budget. Operationally, we adjust all estimates to match the actual mean of the evaluation data (E-data). Details and results on some examples from the data will follow in subsequent drafts.

5 Discussion

The current analysis leaves a few questions to be resolved before proceeding further in addition to the unfinished work described above. First, there are a number of additional exploratory analyses done that were not presented here that have some implications for further modeling work. Also, this version does not explore the impact of the regression to

the mean phenomenon and the potential for hierarchical modeling and so that is discussed briefly below as well.

Exploratory analysis with these models indicated that there were two general problems in the quality of the fit that appeared no matter what model was employed. First, and probably the most problematic, had to do with relatively sick individuals, who had a large number of HCC flags turned on. Our square root link gamma distributed GLM model does well throughout the range of the predictions until we get to the top five or ten percent of the data. Then the model has large and significant negative residuals, indicating that the raw-scale predictions overpay for these cases. If we do the same analysis by the number of HCC's for individual patients, we found that the problem occurs when there are 10 or more HCC flags turned on. The incremental cost of adding one more disease could be less if patients already have several diseases than if they have relatively few or none. Or that incremental cost could be greater, it could be more costly to treat particular diseases if they have other particular comorbidities. Some of this latter effect (e. g. with complications of diabetes) already is built into the HCC structure. In the new Version 6 of the DCG/HCC model, the developers have begun adding interaction terms between HCC's, but only where those incremental costs are significantly positive.

Another way to approach this might be to regroup these cases into a new set of super HCC's - such as complex diabetes with major complications, and turn off the HCC flags for the complication HCC's for these specific diabetic patients. By turning off some of the HCC flags for comorbidities, we avoid this misspecification from assuming a simple additive model. We cannot provide much guidance on how to form these super HCCs. For that, one would need to involve clinicians to put together medically meaningful groups.

The second problem has to do with individuals with no HCC flag versus those with one or more. Exploratory data analysis indicated that there were some problems of fit for those

with no HCC flag when they were combined with the general population. Specifically, the scale parameter was much lower (around 0.08) than for those with any HCC (around 0.7). This implies that distribution for those without any HCC from Table 1 have costs that are far more skewed to the right than those with at least one HCC. Also there is some indication that the coefficients for the demographic variables (age, age squared, female, and age missing) are different. This could suggest employing some type of two part model, though that would be inefficient, or one could deal with this problem more simply by separating the sample into two subpopulations – those with and without any HCCs – without any kind of model for predicting how one would fall in the two groups. We have done the latter here for illustration. For those without, the only covariates would be the demographic variables. For those with any HCC's, we would include HCC, special population flags and demographic variables.

Ongoing work is still pushing to understand the effect of credibility differences across HCC's and the special population flags. The special population flags are of particular interest in the VA with Congressionally mandated attention. We know that the estimates of the special population flags are generally more noisy, reflecting variation in the health care needs and expenditures in their underlying populations, than the estimates of the HCC parameters. The sizes of the number of patients in each HCC and each special population flag group also vary tremendously. Simply from understanding the impact of the regression to the mean inherent in these types of populations, we can infer that hierarchical or multi-level modeling focused on particular flags (or HCC's) of interest would be enlightening. This could be especially important in determining VA VISN allocations when the distribution of these special populations varies greatly by VISN. However, at this point, we do not have models to present in this area, though continuing work is ongoing and shows promise for finding important practical effects.

6 Conclusions

As risk adjustment continues to be a more and more common part of the adjustment processes in resource allocation schemes used by health plans, employers, and governments, this paper illustrates that important distributional aspects of those adjustments may have been overlooked to some extent in previous work. This present paper offers a beginning look at what some of those issues are and suggests this is an important area for further research. In particular, we have suggested, but not studied, the possible role of hierarchical models in focusing attention on the variability in the credibility of mean predictions for particular classes of patients. More generally, we have shown that significant patterns of mispredictions exist in OLS models common in these types of risk adjustment studies. And we have begun to illustrate the issues involved in taking models developed on one sample and applying them to a forecast sample, highlighting the impact of the choice of a loss function. Clearly much more needs to be done to understand these expenditure relationships further and small differences in modeling seemingly can lead to huge differences in resource allocations or prices to providers.

References

[1] Allenby, G.M (1990), "Cross-validation, the Bayes Theorem and Small-Sample Bias", Journal of Business and Economic Statistics, 8, 171-177.

[2] Ash, A, Y. Zhao, R. P. Ellis, and M. S. Kramer (2001), "Finding Future High-cost Cases: Comparing Prior Cost Versus Diagnosis-based Methods," Health Services Research, 36(6) Part II December: 194-206.

[3] Ash A, R. P. Ellis, W. Yu, et al. (1998), "Risk Adjustment for the Non-Elderly." Final Report submitted to the Health Care Financing Administration under cooperative agreement No.18-C-90462/1-02, Boston, MA: Boston University, June 1998.

[4] Ash, A. and S. Byrne-Logan (1998), "How Well do Models Work? Predicting Health Care Costs," in Proceedings of the Section on Statistics in Epidemiology, American Statistical Association, pp. 42-49.

[5] Blough, D.K., C.W. Madden and M.C. Hornbrook (1999), "Modeling Risk using Generalized Linear Models, Journal of Health Economics, 18, 153-171.

[6] DeLong, E.R., E.D. Peterson, D.M. DeLong, L.H. Muhlbaier, S. Hackett, and D.B. Mark, 1997, Comparing risk-adjustment methods for provider profiling, Statistics in Medicine 16, 2645–2664.

[7] Dowd, Feldman, 2001, Some observations about risk adjustment research, Inquiry, 38(3): 315–318.

[8] Duan, N., W.G. Manning, C.N. Morris and J.P. Newhouse (1983), "A Comparison of Alternative Models for the Demand for Medical Care", Journal of Business and Economic Statistics, 1, 115-126.

[9] Duan, N. (1983), "Smearing Estimate: a Nonparametric Retransformation Method", Journal of the American Statistical Association, 78, 605-610.

[10] Dunn, D. (1998), "Applications of Health Risk Adjustments: What Can Be Learned to Date?" Inquiry, 35: 132-147, Summer.

[11] Ellis, R.P., "Creaming, Skimping, and Dumping: Provider Competition on the Intensive and Extensive Margins." Journal of Health Economics, 17:537-555, 1998.

[12] Ellis R. P. and A. Ash (1996), "Refinements to the Diagnostic Cost Group Model," Inquiry, 32: 1-12, Winter.

[13] Ellis, R.P., et al. (1996), "Diagnostic Cost Group (DCG) and Hierarchical Coexisting Condition (HCC) Models for Medicare Risk Adjustment." Final Report to the Health Care Financing Administration, Contract no. 500-92-0020, Waltham, MA: Health Economics Research, Inc, April 1996.

[14] Ellis, R.P., G. Pope, L.I. Iezzoni, J.Z. Ayanian, D.W. Bates, H. Burstin, and A. Ash (1996), "Diagnosis-Based Risk Adjustment for Medicare Capitation Payments." Health Care Financing Review, Spring.

[15] Hosmer, D.W. and S. Lemeshow, Applied Logistic Regression, 2nd Ed., Wiley, (2000).

[16] Jones, A.M., "Health Econometrics", Eds. Culyer, A.J. and Newhouse, J.P. Handbook of Health Economics, Chapter 6:265-344, North Holland, (2000).

[17] Lamers, L.M (1999), "Risk-Adjusted Capitation Based on the Diagnostic Cost Group Model: An Empirical Evaluation with Health Survey Information." Health Services Research, 33(6): 1727-1744, February.

[18] Lehner, L.A., J.F. Burgess, Jr., D. Hults, and T. Stefos, 1996, Data and information requirements for VA resource allocation systems, Medical Care 34, MS21–30.

[19] Manning, W.G. (1998), "The Logged Dependent Variable, Heteroscedasticity, and the Retransformation Problem", Journal of Health Economics, 17, 283-295.

[20] Manning, W.G. and J. Mullahy (2001), "Estimating Log Models: To Transform or Not to Transform?", Journal of Health Economics 20, 461-494.

[21] Mullahy, J. (1998), "Much Ado About Two: Reconsidering Retransformation and the Two-part Model in Health Econometrics", Journal of Health Economics, 17, 247-281.

[22] Newhouse, J. P., "Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection." Journal of Economic Literature, Winter 1995.

[23] Newhouse, J.P., "Risk Adjustment: Where Are We Now?" Inquiry, 35: 122-131, Summer 1998.

[24] Park, R. (1966), "Estimation with Heteroscedastic Error Terms," Econometrica, 34, 888.

[25] Pope, G., R.P. Ellis, C. Liu, et al. (1998), "Revised Diagnostic Cost Group (DCG)/Hierarchical Coexisting Conditions (HCC) Models for Medicare Risk Adjustment." Final Report to the Health Care Financing Administration under Contract No. 500-95-048, Waltham, MA: Health Economics Research, Inc., February 1998.

[26] Rosenberg, C.E., "The Tyranny of Diagnosis: Specific Entities and Individual Experience", The Milbank Quarterly 80, 237–260.

[27] Stefos, T., N. LaVallee, and F. Holden, 1992, Fairness in prospective payment: A clustering approach, Health Services Research 27, 239–261.

[28] van de Ven, W. and R.P. Ellis, "Risk Adjustment in Competitive Health Plan Markets". Eds. Culyer, A.J. and Newhouse, J.P. Handbook of Health Economics, Chapter 14:755-845, North Holland, (2000).

[29] Wilson, V., C. Smith, et al, "Case Study: The Washington State Health Care Authority." Inquiry, 35: 178-192, Summer 1998.

Table 3.
GLM Gamma Distribution with Square Root Link on Full Estimation Samples

Generalized linear models			No. of obs	=	99997
Optimization : ML: Newton-Raphson			Residual df	=	99900
			Scale param		4.091366
Deviance	=	90890.46667	(1/df) Deviance	=	.9098145
Pearson	=	408727.4542	(1/df) Pearson	=	4.091366
Variance function	:	V(u) = u^2	[Gamma]		
Link function	:	g(u) = u^(0.5)	[Power]		
Standard function	:	Sandwich			
Log likelihood	=	-877060.6315	AIC	=	17.54368
BIC	=	-1059247.79			

Parameter	Coefficient	Std. Err.	z	P>|z|	Raw Scale Coefficient
hcc001	27.03761	3.837391	7.05	0.000	$ 731.03
hcc002	37.0489	5.548342	6.68	0.000	$ 1,372.62
hcc003	15.1302	4.265487	3.55	0.000	$ 228.92
hcc004	7.098392	0.314812	22.55	0.000	$ 50.39
hcc005	41.27285	2.45389	16.82	0.000	$ 1,703.45
hcc006	21.90773	1.803466	12.15	0.000	$ 479.95
hcc007	16.98955	1.801931	9.43	0.000	$ 288.64
hcc008	11.8947	0.5218085	22.80	0.000	$ 141.48
hcc013	12.2983	0.572718	21.47	0.000	$ 151.25
hcc014	13.59827	0.5374534	25.30	0.000	$ 184.91
hcc015	8.511202	0.2608323	32.63	0.000	$ 72.44
hcc016	29.28046	5.263281	5.56	0.000	$ 857.35
hcc017	15.13539	1.186152	12.76	0.000	$ 229.08
hcc019	10.51755	2.329664	4.51	0.000	$ 110.62
hcc020	13.4038	1.218009	11.00	0.000	$ 179.66
hcc021	32.78456	2.202588	14.88	0.000	$ 1,074.83
hcc022	10.26744	0.6207715	16.54	0.000	$ 105.42
hcc023	6.958924	0.2406935	28.91	0.000	$ 48.43
hcc024	22.72831	2.861858	7.94	0.000	$ 516.58
hcc025	10.03982	0.7215449	13.91	0.000	$ 100.80
hcc027	16.97525	3.043503	5.58	0.000	$ 288.16
hcc028	7.997845	1.287043	6.21	0.000	$ 63.97
hcc029	9.107978	0.5607819	16.24	0.000	$ 82.96
hcc030	13.52238	3.110096	4.35	0.000	$ 182.85
hcc031	33.16345	0.7372553	44.98	0.000	$ 1,099.81
hcc032	24.41109	0.5315584	45.92	0.000	$ 595.90
hcc033	10.99506	0.3397994	32.36	0.000	$ 120.89
hcc040	56.57326	13.02074	4.34	0.000	$ 3,200.53

hcc041	36.67856	12.40745	2.96	0.003	$ 1,345.32
hcc042	10.5221	0.8140749	12.93	0.000	$ 110.71
hcc043	12.05017	0.6348274	18.98	0.000	$ 145.21
hcc044	8.200676	0.4690315	17.48	0.000	$ 67.25
hcc045	74.80313	14.93662	5.01	0.000	$ 5,595.51
hcc046	19.9368	5.005543	3.98	0.000	$ 397.48
hcc047	31.70091	4.277416	7.41	0.000	$ 1,004.95
hcc048	8.910051	0.5512447	16.16	0.000	$ 79.39
hcc049	8.308866	0.5309679	15.65	0.000	$ 69.04
hcc050	55.50527	2.657818	20.88	0.000	$ 3,080.83
hcc051	34.16908	1.532518	22.30	0.000	$ 1,167.53
hcc053	8.32644	0.7997688	10.41	0.000	$ 69.33
hcc054	7.719745	1.106672	6.98	0.000	$ 59.59
hcc055	10.63898	2.373941	4.48	0.000	$ 113.19
hcc058	12.89416	2.12436	6.07	0.000	$ 166.26
hcc059	9.035481	0.6682337	13.52	0.000	$ 81.64
hcc060	9.115826	0.5796742	15.73	0.000	$ 83.10
hcc061	17.49071	1.822838	9.60	0.000	$ 305.92
hcc063	7.549301	1.251873	6.03	0.000	$ 56.99
hcc064	6.832491	0.2879035	23.73	0.000	$ 46.68
hcc065	68.67059	6.612438	10.39	0.000	$ 4,715.65
hcc066	33.56563	5.38893	6.23	0.000	$ 1,126.65
hcc067	14.97386	1.221024	12.26	0.000	$ 224.22
hcc068	5.099481	1.303214	3.91	0.000	$ 26.00
hcc069	29.23801	4.033903	7.25	0.000	$ 854.86
hcc075	5.826871	0.2141078	27.21	0.000	$ 33.95
hcc076	13.60724	4.000321	3.40	0.001	$ 185.16
hcc077	35.1256	3.22401	10.90	0.000	$ 1,233.81
hcc078	11.97514	1.131379	10.58	0.000	$ 143.40
hcc079	5.221548	1.593386	3.28	0.001	$ 27.26
hcc080	9.346125	0.5002736	18.68	0.000	$ 87.35
hcc091	16.54201	1.902089	8.70	0.000	$ 273.64
hcc093	15.88933	3.620986	4.39	0.000	$ 252.47
hcc094	25.86051	4.171899	6.20	0.000	$ 668.77
hcc095	17.03942	4.353961	3.91	0.000	$ 290.34
hcc096	13.26853	2.007342	6.61	0.000	$ 176.05
hcc097	9.119974	0.3826319	23.83	0.000	$ 83.17
hcc098	41.66444	2.202886	18.91	0.000	$ 1,735.93
hcc099	9.670223	0.3129953	30.90	0.000	$ 93.51
hcc100	7.391308	0.2168686	34.08	0.000	$ 54.63
hcc102	19.27662	5.467267	3.53	0.000	$ 371.59
hcc103	8.987262	1.330211	6.76	0.000	$ 80.77
hcc110	33.4732	4.832966	6.93	0.000	$ 1,120.46
hcc111	12.582	3.956808	3.18	0.001	$ 158.31
hcc112	19.96072	3.751275	5.32	0.000	$ 398.43
hcc113	10.50856	0.3435536	30.59	0.000	$ 110.43
hcc114	18.72861	4.570601	4.10	0.000	$ 350.76

hcc115	27.59942	3.683949	7.49	0.000	$ 761.73
hcc116	20.45333	0.9766179	20.94	0.000	$ 418.34
hcc118	6.524408	0.4717504	13.83	0.000	$ 42.57
flag1	33.00486	10.21412	3.23	0.001	$ 1,089.32
flag2	132.5377	4.93026	26.88	0.000	$17,566.24
flag3	40.2808	17.08991	2.36	0.018	$ 1,622.54
flag4	105.2284	5.830672	18.05	0.000	$11,073.02
flag5	51.19106	3.482031	14.70	0.000	$ 2,620.52
flag6	106.1185	9.272044	11.44	0.000	$11,261.14
flag7	62.33075	5.86182	10.63	0.000	$ 3,885.12
flag8	35.18357	3.854876	9.13	0.000	$ 1,237.88
flag9	23.35973	2.866188	8.15	0.000	$ 545.68
flag10	62.71267	2.728571	22.98	0.000	$ 3,932.88
flag11	136.3633	3.671568	37.14	0.000	$18,594.95
flag12	32.88679	4.350088	7.56	0.000	$ 1,081.54
flag13	97.94384	23.79594	4.12	0.000	$ 9,593.00
flag14	118.6732	6.781541	17.50	0.000	$14,083.33
agemiss	-0.6471098	2.344377	-0.28	0.783	
age	0.0430026	0.0063162	6.81	0.000	
age2	-0.0056095	0.0003501	-16.02	0.000	
female	2.109217	0.3657939	5.77	0.000	$ 4.45
_cons	25.31801	0.2279175	111.08	0.000	$ 641.00

Table 3A.
OLS Regression Results with Intercept on Full Estimation Sample

Number of ob	=	99997
F(96, 99900)	=	159.95
Prob > F	=	0.0000
R-squared	=	0.4871
Root MSE	=	10697

Parameter	Coefficient	Std. Err.	t	P>|t|
hcc001	4088.208	1039.468	3.93	0.000
hcc002	14913.68	2929.998	5.09	0.000
hcc003	3010.182	2101.459	1.43	0.152
hcc004	1234.724	145.24	8.50	0.000
hcc005	6619.386	931.7445	7.10	0.000
hcc006	3037.228	673.9495	4.51	0.000
hcc007	2656.679	563.9675	4.71	0.000
hcc008	1199.196	213.8971	5.61	0.000
hcc013	428.1034	334.9596	1.28	0.201
hcc014	866.2725	225.0498	3.85	0.000
hcc015	671.197	115.0851	5.83	0.000
hcc016	7214.065	1510.21	4.78	0.000
hcc017	4605.538	537.7566	8.56	0.000
hcc019	1167.614	616.7736	1.89	0.058
hcc020	1525.963	536.6976	2.84	0.004
hcc021	7145.931	803.5789	8.89	0.000
hcc022	2221.21	259.1573	8.57	0.000
hcc023	406.0969	98.46357	4.12	0.000
hcc024	8193.3	1206.332	6.79	0.000
hcc025	513.9542	255.1861	2.01	0.044
hcc027	4027.672	1462.959	2.75	0.006
hcc028	3274.287	666.6331	4.91	0.000
hcc029	1412.806	266.7465	5.30	0.000
hcc030	2696.804	444.2445	6.07	0.000
hcc031	3990.833	198.4598	20.11	0.000
hcc032	2431.643	146.6024	16.59	0.000
hcc033	884.3966	152.2694	5.81	0.000
hcc040	13605.53	3882.012	3.50	0.000
hcc041	10848.61	2751.805	3.94	0.000
hcc042	1737.98	469.471	3.70	0.000
hcc043	1825.659	239.6132	7.62	0.000
hcc044	928.5503	178.3328	5.21	0.000
hcc045	41092.9	7534.905	5.45	0.000
hcc046	2785.452	5128.369	0.54	0.587
hcc047	12279.57	1531.755	8.02	0.000
hcc048	1256.059	242.5945	5.18	0.000
hcc049	1691.019	288.7089	5.86	0.000

hcc050	10879.92	1132.135	9.61	0.000
hcc051	6755.445	495.9236	13.62	0.000
hcc053	1629.538	376.2969	4.33	0.000
hcc054	907.2913	636.5149	1.43	0.154
hcc055	1181.023	617.1636	1.91	0.056
hcc058	4490.26	809.6868	5.55	0.000
hcc059	607.1261	248.7473	2.44	0.015
hcc060	1890.748	250.8106	7.54	0.000
hcc061	6713.511	1044.54	6.43	0.000
hcc063	1141.744	372.8471	3.06	0.002
hcc064	620.4885	143.2716	4.33	0.000
hcc065	24870.19	2864.58	8.68	0.000
hcc066	8711.527	1793.408	4.86	0.000
hcc067	5284.812	733.5532	7.20	0.000
hcc068	1105.236	776.2217	1.42	0.154
hcc069	12207.88	2014.409	6.06	0.000
hcc075	235.6607	92.45211	2.55	0.011
hcc076	10205.33	3271.114	3.12	0.002
hcc077	2900.954	1172.107	2.47	0.013
hcc078	3509.624	571.0071	6.15	0.000
hcc079	1266.218	841.9265	1.50	0.133
hcc080	1903.54	220.2423	8.64	0.000
hcc091	8582.05	900.1615	9.53	0.000
hcc093	1603.976	1857.121	0.86	0.388
hcc094	8236.764	1820.002	4.53	0.000
hcc095	2166.992	1647.155	1.32	0.188
hcc096	3428	864.4928	3.97	0.000
hcc097	1055.515	155.1329	6.80	0.000
hcc098	14718.53	898.8628	16.37	0.000
hcc099	950.5977	126.3074	7.53	0.000
hcc100	480.7236	82.744	5.81	0.000
hcc102	10279.75	5342.868	1.92	0.054
hcc103	846.4145	415.5246	2.04	0.042
hcc110	5120.287	2492.157	2.05	0.040
hcc111	2833.257	1979.356	1.43	0.152
hcc112	6202.884	1386.714	4.47	0.000
hcc113	1109.44	143.1452	7.75	0.000
hcc114	6135.369	2718.433	2.26	0.024
hcc115	3632.744	1411.392	2.57	0.010
hcc116	5378.535	431.3861	12.47	0.000
hcc118	1454.657	218.2523	6.67	0.000
flag1	8508.872	2313.342	3.68	0.000
flag2	38033.54	1899.655	20.02	0.000
flag3	10508.23	3461.454	3.04	0.002
flag4	22485.4	1952.906	11.51	0.000
flag5	11994.16	1059.259	11.32	0.000
flag6	32807.07	3405.18	9.63	0.000

flag7	11438.6	1360.284	8.41	0.000
flag8	3271.527	1081.399	3.03	0.002
flag9	3344.31	888.6512	3.76	0.000
flag10	14517.2	814.1472	17.83	0.000
flag11	38074.04	1557.223	24.45	0.000
flag12	7332.197	1262.831	5.81	0.000
flag13	58074.17	18013.61	3.22	0.001
flag14	38192.41	3161.251	12.08	0.000
agemiss	-144.1105	123.2517	-1.17	0.242
age	-17.41363	2.493696	-6.98	0.000
age2	-1.23094	0.1358774	-9.06	0.000
female	562.9697	143.5453	3.92	0.000
_cons	-2.084651	54.74119	-0.04	0.970

Table 4.
GLM Gamma Distribution with Square Root Link on Estimation Sample with at least One HCC

Generalized linear models		No. of obs	= 81725
Optimization : ML: Newton-Raphson		Residual df	= 81628
		Scale param	1.464479
Deviance	= 63308.9103	(1/df) Deviance	= .7755784
Pearson	= 119542.5102	(1/df) Pearson	= 1.464479
Variance function	: V(u) = u^2	[Gamma]	
Link function	: g(u) = u^(0.5)	[Power]	
Standard function	: Sandwich		
Log likelihood	= -742347.4973	AIC	= 18.16934
BIC	= -859994.8038		

Parameter	Coefficient	Std. Err.	z	P>|z|	Raw Scale Coefficient
hcc001	25.27114	3.499906	7.22	0.000	$ 638.63
hcc002	36.8054	5.522097	6.67	0.000	$ 1,354.64
hcc003	14.98808	4.267055	3.51	0.000	$ 224.64
hcc004	7.095097	0.310089	22.88	0.000	$ 50.34
hcc005	41.26798	2.434179	16.95	0.000	$ 1,703.05
hcc006	21.8738	1.793386	12.20	0.000	$ 478.46
hcc007	16.91033	1.767033	9.57	0.000	$ 285.96
hcc008	12.1049	0.51184	23.65	0.000	$ 146.53
hcc013	12.3473	0.59755	20.66	0.000	$ 152.46
hcc014	13.44546	0.536495	25.06	0.000	$ 180.78
hcc015	8.551488	0.27181	31.46	0.000	$ 73.13
hcc016	28.95015	5.169609	5.60	0.000	$ 838.11
hcc017	15.13357	1.182988	12.79	0.000	$ 229.02
hcc019	10.24351	2.285633	4.48	0.000	$ 104.93
hcc020	13.38187	1.237136	10.82	0.000	$ 179.07
hcc021	32.77664	2.199986	14.90	0.000	$ 1,074.31
hcc022	10.21948	0.614272	16.64	0.000	$ 104.44
hcc023	6.976457	0.241074	28.94	0.000	$ 48.67
hcc024	22.72199	2.813077	8.08	0.000	$ 516.29
hcc025	10.16063	0.756178	13.44	0.000	$ 103.24
hcc027	17.17586	2.960106	5.80	0.000	$ 295.01
hcc028	7.865595	1.254881	6.27	0.000	$ 61.87
hcc029	9.335093	0.557722	16.74	0.000	$ 87.14
hcc030	13.80784	2.9022	4.76	0.000	$ 190.66
hcc031	32.75761	0.740461	44.24	0.000	$ 1,073.06
hcc032	23.99099	0.53378	44.95	0.000	$ 575.57
hcc033	10.73679	0.345485	31.08	0.000	$ 115.28

hcc040	46.54109	11.34934	4.10	0.000	$ 2,166.07
hcc041	26.77885	8.74582	3.06	0.002	$ 717.11
hcc042	10.3392	0.817386	12.65	0.000	$ 106.90
hcc043	12.03631	0.625117	19.25	0.000	$ 144.87
hcc044	8.127387	0.468809	17.34	0.000	$ 66.05
hcc045	74.44934	14.8569	5.01	0.000	$ 5,542.70
hcc046	19.79314	5.123981	3.86	0.000	$ 391.77
hcc047	31.71795	4.361978	7.27	0.000	$ 1,006.03
hcc048	9.108757	0.562201	16.20	0.000	$ 82.97
hcc049	8.617107	0.526931	16.35	0.000	$ 74.25
hcc050	55.40124	2.658954	20.84	0.000	$ 3,069.30
hcc051	33.86877	1.531791	22.11	0.000	$ 1,147.09
hcc053	8.395578	0.787273	10.66	0.000	$ 70.49
hcc054	7.991985	1.083512	7.38	0.000	$ 63.87
hcc055	10.86332	2.308382	4.71	0.000	$ 118.01
hcc058	12.91116	2.13075	6.06	0.000	$ 166.70
hcc059	9.114973	0.656019	13.89	0.000	$ 83.08
hcc060	9.221782	0.580406	15.89	0.000	$ 85.04
hcc061	17.2653	1.811722	9.53	0.000	$ 298.09
hcc063	7.493298	1.185422	6.32	0.000	$ 56.15
hcc064	7.004521	0.286591	24.44	0.000	$ 49.06
hcc065	68.30036	6.636715	10.29	0.000	$ 4,664.94
hcc066	33.51266	5.376845	6.23	0.000	$ 1,123.10
hcc067	14.84611	1.219621	12.17	0.000	$ 220.41
hcc068	4.976438	1.285976	3.87	0.000	$ 24.76
hcc069	29.19102	4.043939	7.22	0.000	$ 852.12
hcc075	5.802032	0.224696	25.82	0.000	$ 33.66
hcc076	13.13951	3.940456	3.33	0.001	$ 172.65
hcc077	34.72456	3.219184	10.79	0.000	$ 1,205.80
hcc078	11.97124	1.109653	10.79	0.000	$ 143.31
hcc079	5.713827	1.62924	3.51	0.000	$ 32.65
hcc080	9.325042	0.499009	18.69	0.000	$ 86.96
hcc091	16.82524	2.172975	7.74	0.000	$ 283.09
hcc093	15.53241	3.563022	4.36	0.000	$ 241.26
hcc094	26.0374	4.182908	6.22	0.000	$ 677.95
hcc095	16.97976	4.259469	3.99	0.000	$ 288.31
hcc096	13.09687	2.003673	6.54	0.000	$ 171.53
hcc097	9.046525	0.372225	24.30	0.000	$ 81.84
hcc098	41.42074	2.2177	18.68	0.000	$ 1,715.68
hcc099	9.62914	0.311087	30.95	0.000	$ 92.72
hcc100	7.396169	0.221839	33.34	0.000	$ 54.70
hcc102	19.07271	5.423632	3.52	0.000	$ 363.77
hcc103	8.842293	1.294599	6.83	0.000	$ 78.19
hcc110	33.00696	4.820913	6.85	0.000	$ 1,089.46
hcc111	13.34046	3.850443	3.46	0.001	$ 177.97
hcc112	19.83678	3.731595	5.32	0.000	$ 393.50
hcc113	10.59049	0.344011	30.79	0.000	$ 112.16

hcc114	18.6463	4.595802	4.06	0.000	$ 347.68
hcc115	27.56324	3.635873	7.58	0.000	$ 759.73
hcc116	20.27509	0.978422	20.72	0.000	$ 411.08
hcc118	6.533735	0.466412	14.01	0.000	$ 42.69
flag1	44.83911	9.502568	4.72	0.000	$ 2,010.55
flag2	132.607	4.923589	26.93	0.000	$17,584.62
flag3	45.32113	16.39869	2.76	0.006	$ 2,054.00
flag4	105.6444	5.833802	18.11	0.000	$11,160.74
flag5	51.00868	3.486888	14.63	0.000	$ 2,601.89
flag6	106.0025	9.228555	11.49	0.000	$11,236.53
flag7	63.51704	5.908269	10.75	0.000	$ 4,034.41
flag8	34.85903	3.833618	9.09	0.000	$ 1,215.15
flag9	25.05099	3.040064	8.24	0.000	$ 627.55
flag10	62.67373	2.717238	23.07	0.000	$ 3,928.00
flag11	137.465	3.689668	37.26	0.000	$18,896.63
flag12	35.34776	3.936275	8.98	0.000	$ 1,249.46
flag13	97.80717	23.82512	4.11	0.000	$ 9,566.24
flag14	120.0965	6.834461	17.57	0.000	$14,423.17
agemiss	-10.25863	0.533396	-19.23	0.000	
age	0.0127221	0.007233	1.76	0.079	
age2	-0.0076114	0.000391	-19.47	0.000	
female	2.844564	0.499719	5.69	0.000	$ 8.09
_cons	25.90467	0.245964	105.32	0.000	$ 671.05

Table 4A.
OLS Regression Results with Intercept on Estimation Sample with at least One HCC

Number of obs	=	81725
F (96, 81628)	=	136.76
Prob > F	=	0.0000
R-squared	=	0.4790
Root MSE	=	11759

Parameter	Coefficient	Std. Err.	t	P>|t|
hcc001	3987.615	1065.922	3.74	0.000
hcc002	14819.17	2925.648	5.07	0.000
hcc003	2996.6	2103.566	1.42	0.154
hcc004	1317.073	145.7769	9.03	0.000
hcc005	6703.065	931.2422	7.20	0.000
hcc006	3173.456	673.8521	4.71	0.000
hcc007	2814.141	563.8228	4.99	0.000
hcc008	1420.493	215.2098	6.60	0.000
hcc013	548.409	335.1567	1.64	0.102
hcc014	1111.958	226.6534	4.91	0.000
hcc015	953.6946	119.7061	7.97	0.000
hcc016	7155.309	1510.056	4.74	0.000
hcc017	4534.916	537.1255	8.44	0.000
hcc019	1159.951	616.4258	1.88	0.060
hcc020	1658.073	536.3832	3.09	0.002
hcc021	7186.219	803.0666	8.95	0.000
hcc022	2369.001	260.3015	9.10	0.000
hcc023	608.9742	100.5685	6.06	0.000
hcc024	8204.538	1205.179	6.81	0.000
hcc025	648.1215	255.6645	2.54	0.011
hcc027	4038.788	1461.079	2.76	0.006
hcc028	3301.461	665.7567	4.96	0.000
hcc029	1472.186	266.6881	5.52	0.000
hcc030	2783.606	445.815	6.24	0.000
hcc031	4180.857	200.8376	20.82	0.000
hcc032	2620.128	149.501	17.53	0.000
hcc033	1088.222	155.1832	7.01	0.000
hcc040	13171.41	3959.362	3.33	0.001
hcc041	10577.68	2770.463	3.82	0.000
hcc042	1747.258	469.3585	3.72	0.000
hcc043	1947.188	240.0853	8.11	0.000
hcc044	1030.305	178.4948	5.77	0.000
hcc045	40962.09	7539.159	5.43	0.000
hcc046	2793.457	5110.768	0.55	0.585
hcc047	12220.55	1530.548	7.98	0.000
hcc048	1321.59	242.6809	5.45	0.000
hcc049	1815.441	289.5668	6.27	0.000

hcc050	10815.42	1131.731	9.56	0.000
hcc051	6691.683	495.0347	13.52	0.000
hcc053	1709.085	376.2789	4.54	0.000
hcc054	960.0559	636.336	1.51	0.131
hcc055	1335.795	615.532	2.17	0.030
hcc058	4488.059	810.3	5.54	0.000
hcc059	709.8355	249.2136	2.85	0.004
hcc060	1992.198	251.0958	7.93	0.000
hcc061	6681.529	1044.329	6.40	0.000
hcc063	1217.621	372.8804	3.27	0.001
hcc064	793.078	145.2521	5.46	0.000
hcc065	24737.02	2864.823	8.63	0.000
hcc066	8569.734	1789.146	4.79	0.000
hcc067	5226.349	733.1601	7.13	0.000
hcc068	1141.136	776.8036	1.47	0.142
hcc069	12114.56	2011.298	6.02	0.000
hcc075	417.2737	95.42828	4.37	0.000
hcc076	10096.49	3266.782	3.09	0.002
hcc077	3097.293	1166.843	2.65	0.008
hcc078	3560.471	570.8714	6.24	0.000
hcc079	1319.339	841.2959	1.57	0.117
hcc080	1987.999	220.2686	9.03	0.000
hcc091	8518.678	900.1859	9.46	0.000
hcc093	1564.828	1856.261	0.84	0.399
hcc094	8304.122	1818.235	4.57	0.000
hcc095	2091.366	1657.321	1.26	0.207
hcc096	3440.557	862.8264	3.99	0.000
hcc097	1165.056	156.3706	7.45	0.000
hcc098	14618.89	897.6149	16.29	0.000
hcc099	994.6197	126.6107	7.86	0.000
hcc100	678.9651	86.21254	7.88	0.000
hcc102	10304.46	5331.741	1.93	0.053
hcc103	896.591	415.6599	2.16	0.031
hcc110	5213.057	2490.415	2.09	0.036
hcc111	2929.49	1975.747	1.48	0.138
hcc112	6211.789	1386.403	4.48	0.000
hcc113	1220.903	144.0509	8.48	0.000
hcc114	6038.466	2719.358	2.22	0.026
hcc115	3542.115	1411.693	2.51	0.012
hcc116	5341.255	431.1433	12.39	0.000
hcc118	1515.192	218.4834	6.94	0.000
flag1	9098.852	2466.97	3.69	0.000
flag2	37994.52	1901.11	19.99	0.000
flag3	11115.85	3616.641	3.07	0.002
flag4	22713.36	1948.414	11.66	0.000
flag5	11827.05	1057.529	11.18	0.000
flag6	32598.11	3405.566	9.57	0.000

flag7	11617.48	1381.682	8.41	0.000
flag8	3214.391	1081.827	2.97	0.003
flag9	3325.122	920.6675	3.61	0.000
flag10	14395.75	812.7865	17.71	0.000
flag11	38443.23	1581.178	24.31	0.000
flag12	7665.315	1312.36	5.84	0.000
flag13	57785.96	17994.52	3.21	0.001
flag14	38573.49	3191.675	12.09	0.000
agemiss	-823.7293	160.157	-5.14	0.000
age	-22.19263	3.219083	-6.89	0.000
age2	-1.486504	0.1733101	-8.58	0.000
female	644.7945	177.9571	3.62	0.000
_cons	-552.5609	85.10637	-6.49	0.000

Figure 1: Full Sample Modified Hosmer/Lemeshow Test Results

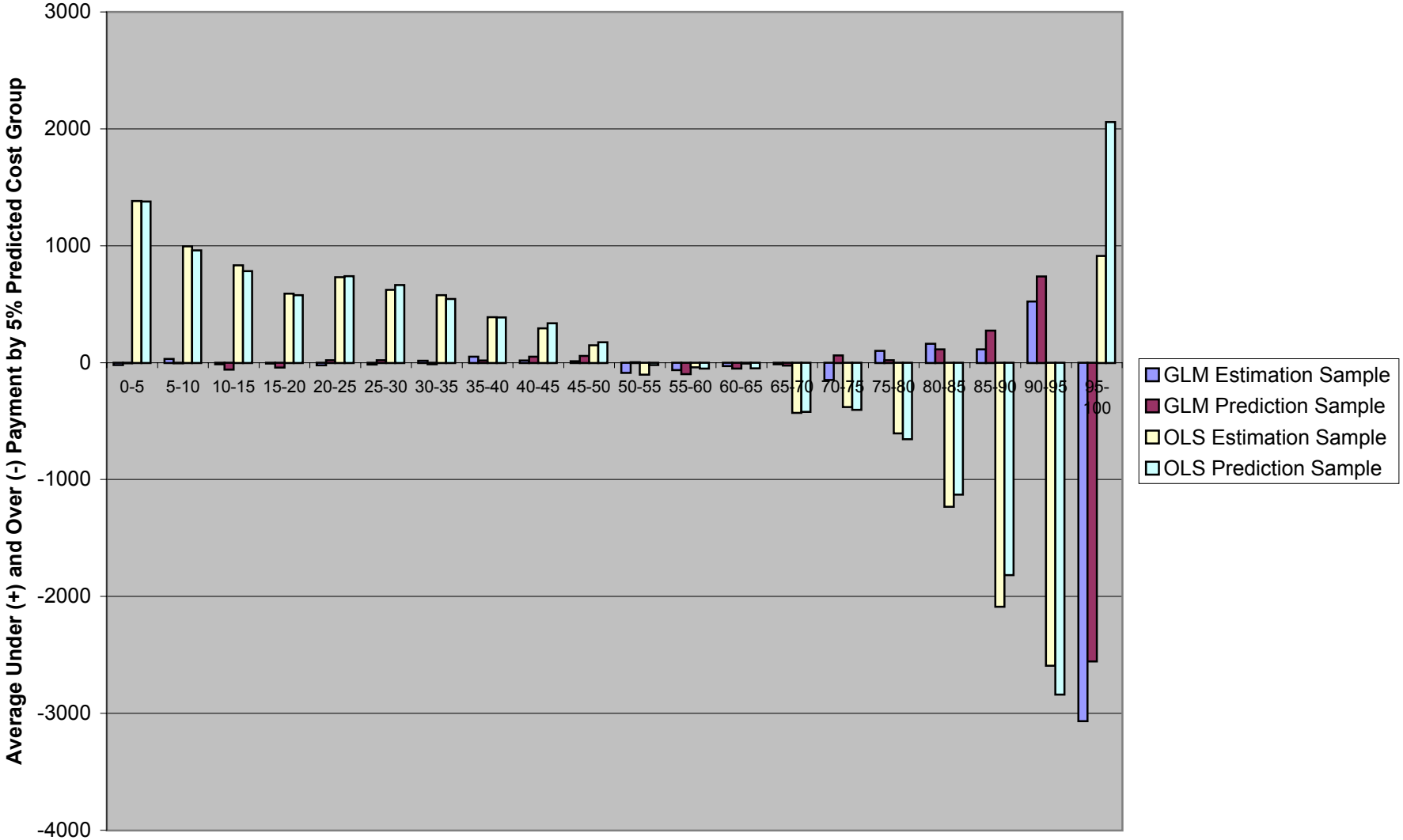

Figure 2: Restricted Sample (at least one HCC) Modified Hosmer/Lemeshow Test Results